

# Інформаційна технологія аналізу та синтезу гістологічних зображень

*Олег Березький, Мельник Григорій*

Кафедра комп'ютерної інженерії  
Тернопільський національний економічний університет, Тернопіль  
mgm@tneu.edu.ua, ob@tneu.edu.ua

## Анотація

Описано інформаційну технологію аналізу гістологічних зображень. Сучасні технології направлені на детекцію та аналіз окремих мікрооб'єктів, що робить актуальним розроблення алгоритмічного забезпечення для оцінки стану тканин та інших структурних утворень. В основі технології лежать структурні методи аналізу зображень. Для оцінки структурних змін в тканинах органів запропоновано метод на основі застосування груп симетрії, що дозволяє отримати інваріантний до повороту, зсуву та масштабу структурний опис зображень.

## 1. Вступ

Рак є однією з найпоширеніших причин смерті на планеті і за даними ВОЗ до 2030 року глобальна смертність від нього у світі зросте порівняно з 2007 роком у півтора рази і складе 11,5 мільйонів випадків. Діагностування ракових захворювань зазвичай вимагає гістопатологічного дослідження, тобто дослідження пошкоджених тканин людини. Гістологія - це наука, що вивчає закономірності розвитку, будови і функції тканин, а також міжтканинної взаємодії. Об'єкт гістології - тканини - є топологічно і функціонально зв'язані клітинні системи і їх похідні, з яких утворені органи.

Основним методом гістологічного дослідження клітин, тканин і органів є світлова мікроскопія, яка є джерелом гістологічних (ГЗ) та цитологічних (ЦЗ) зображень. Мікрооб'єкти на ГЗ і ЦЗ це клітини, що складаються з ядра, оточеного цитоплазмою. Різні класи клітин або стани однієї клітини можуть відрізнятися як наявністю цитоплазми взагалі, так і формою та кількістю ядер. Структурами на ГЗ є складні об'єкти, що утворені клітинами: шари клітин, стінки судин, залози та ін.

У клінічній практиці при дослідженні злоякісних новоутворень та постановці діагнозу використовується світлова мікроскопія, яка є областю лабораторної діагностики, де домінує трудомісткий суб'єктивний якісний аналіз.

Для автоматизації мікроскопічних досліджень застосовуються системи автоматизованої мікроскопії (САМ). До їх складу входить спеціалізоване програмне і апаратне забезпечення. Актуальність їх розроблення обґрунтована вдосконаленням якості морфологічної діагностики.

Для аналізу ЦЗ та ГЗ на ринку існує широкий ряд програмних засобів та програмно-апаратних комплексів [1]. Всі вони відрізняються по функціональних можливостях, спеціалізації, архітектурі, рівню автоматизації процесу аналізу.

По функціональних можливостях системи можна розділити на такі типи:

- системи отримання зображення із фотокамери, відеокамери, або цифрового мікроскопа;
- системи, що містять засоби для ручного інтерактивного вимірювання розмірів мікрооб'єктів та структур, підрахунку їх кількості та статистичного аналізу;
- системи які мають засоби для автоматизованого виділення мікрооб'єктів шляхом підбору параметрів операцій сегментації;
- системи, що містять засоби класифікації зображення, мікрооб'єктів чи станів мікрооб'єкта;
- експертні системи, що дозволяють будувати якісний опис станів мікрооб'єктів.

Рівень функціональності системи та її призначення впливають на її програмну архітектуру. Можна виділити три підходи до побудови програмних засобів систем аналізу ГЗ. Перший орієнтований на роботу оператора з одним поточним зображенням над яким виконуються операції, що викликаються із меню. Другий орієнтований на оброблення серії зображень з допомогою окремих модулів (підпрограм) кожен з яких виконує певну операцію. Третій підхід полягає в поєднанні попередніх двох.

З точки зору рівня аналізу зображення системи оброблення ГЗ та ЦЗ можна умовно розділити на три класи. Зазначимо, що засоби для проведення морфометричних вимірювань та підрахунку об'єктів, як правило, присутні у всіх системах. Прості редактори обмежені функціями отримання зображення та ручним вимірюванням лінійних розмірів (ScreenMeter, Motic Images Advanced 3.2).

Системи оброблення ГЗ містять операції детекції клітин та їх класифікації (ВидеоТесТ-Морфологія 5.0, AnalySIS Five, BioVision, QCapture PRO 6.0, CellProfiler).

Експертні системи системи аналізу зображень (ЕСАЗ) використовують знання експертів патологів для опису мікрооб'єктів, характеристик, патологічних процесів у якісних категоріях і висновку про стан злоякісного процесу (KIDS).

Системи аналізу ГЗ також можна розділити за призначенням (задачею) і відповідним типом результуючої інформації:

- ручне виділення мікрооб'єктів, підрахунок їх кількості, вимірювання числових ознак, статистичне оброблення результатів;

- автоматичне виділення (класифікація) мікрооб'єктів, підрахунок їх кількості, визначення числових ознак, статистичне оброблення результатів;
- вивід висновку про діагноз на основі кількості об'єктів певного класу (цитологія).

Проте для частини діагностичних процесів, що базуються на структурних характеристиках тканин і взаємному положенні клітин математичне та алгоритмічне забезпечення все ще відсутнє. Тому актуальною задачею є розроблення методів і алгоритмів структурного аналізу гістологічних зображень і побудова на їх основі інформаційної технології.

## 2. Інформаційні технології опрацювання гістологічних зображень

Інформаційна технологія - це системно-організована послідовність операцій, що виконуються над інформацією з використанням засобів і методів автоматизації [2, 3]. Операціями є елементарні дії над інформацією. Інформаційна технологія опрацювання зображення розглядається як послідовність операцій по його перетворенню у потрібну форму.

Хоча сама технологія залежить від виду вирішуваної задачі вона завжди містить операції попередньої обробки для покращення зображення, операції отримання формалізованого представлення зображення (бінаризація, виділення контурів, областей із заданими характеристиками та примітивів та ін.). Процес аналізу виділених примітивів об'єктів часто є заключним етапом технології опрацювання зображень [2].

Інформаційним забезпечення технології є наступне:

1. направлення на гістологічне дослідження;
2. серію зображень, які відносяться до одного експерименту, зразка або препарату;
3. звіт патолога про гістологічне дослідження.

Узагальнену діаграму технології оброблення ГЗ та ЦЗ в САМ можна представити послідовністю операцій показаною на рисунку 1. Слід відмітити, що різні системи залежно від рівня складності містять не всі перелічені операції. Більшість САМ обмежуються визначенням числових ознак мікрооб'єктів. Складніші інтелектуальні системи, що засновані на знаннях, визначають тип злоякісного новоутворення (в інтерактивному або автоматизованому режимах).

Отримання цифрового зображення виконується з допомогою цифрового мікроскопа, фотокамери або слайд-сканера. Операція попереднього оброблення застосовується для усунення дефектів, покращення якості, корекції яскравості та ін. Наступним кроком є детекція простих мікрооб'єктів (клітин) з допомогою різних алгоритмів сегментації в ручному або автоматизованому режимах. Після виділення мікрооб'єктів обов'язковим етапом є обчислення їх геометричних ознак які використовуються для статистичного аналізу. Для подальшого аналізу застосовують ЕСАЗ, що призначені для вирішення задач класифікації зображень, розпізнавання об'єктів та виводу висновку про стан об'єктаів. ЕСАЗ широко використовують наступні методи штучного інтелекту: якісні міркування, нечітку логіку, штучні нейронні мережі.

Слід перелічити ряд проблем, що утруднюють розроблення та впровадження інформаційних технологій в процес діагностики на основі ГЗ.

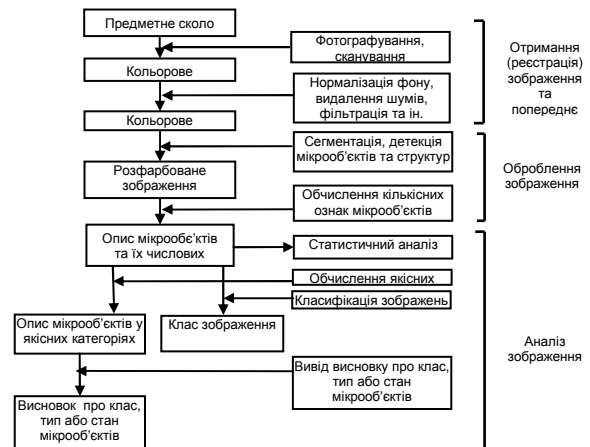


Рисунок 1: Діаграма технології оброблення ГЗ

Більшість засобів орієнтовані на знання користувачем принципів та методів оброблення зображень. Це відображено в тому, що ГЗ, об'єкти на ньому та дії над зображенням описані в термінах комп'ютерного зору. Ефективне використання такого ПЗ можливе тільки у випадку коли оператор відповідну технічну освіту або спеціальну підготовку.

Мінімізація рутинної роботи під час введення ГЗ забезпечується:

- автоматичним вибором інформативних полів зору на основі оцінки локальних і глобальних ознак зображень;
- застосуванням моторизованих мікроскопів, що дозволяє сканувати предметне скло автоматично та отримувати зображення всього зразку у вигляді віртуального слайду.
- застосуванням засобів автоматичної подачі зразків на предметний стіл мікроскопа.

Іншою проблемою є використання інтерактивних алгоритмів, що потребують підстроювання їх параметрів. Результат роботи алгоритму при цьому контролюється користувачем візуально. Послідовність застосування алгоритмів та операторів оброблення зображення також задається вручну. Такі алгоритми повинні виконуватися в інтерактивному режимі:

- для розв'язання проблеми надійності: щоб гарантувати, що користувач повністю контролює результати виконання алгоритму шляхом інтерактивного управління;
- для розв'язання питань правової відповідальності: автоматична обробка медичних даних часто створює проблему правової відповідальності. Якщо алгоритм виконується під керівництвом користувача, ця проблема знімається.

### 3. Методи, алгоритми аналізу і синтезу гістологічних зображень

#### 3.1. Пошук зображень в базі даних

Одним з найбільш важливих аспектів при побудові великомасштабних віртуальних наборів даних для зображень є можливість виконувати запити користувача ефективним способом. Для забезпечення пошуку об'єктів на ГЗ розроблено алгоритмічне забезпечення [4], що використовує ознаки форми та кольору мікрооб'єктів.

Колірна ознака зображення клітин формується як набір значень кольору кожного кластера і його відносного розміру. Введемо позначення:  $A$  - перше зображення,  $K^A$  - кількість кластерів для зображення  $A$ ,  $I_i$  - значення кольору  $i$ -го кластера,  $S_i$  - відносний розмір  $i$ -го кластера. Колірна ознака  $f_A$  для зображення  $A$  визначається так:

$$f_A = \{(I_i, S_i) \mid I_i \in \{R, G, B\}, \\ R, G, B \in \{0, 1, \dots, 255\}, 0 \leq S_i \leq 1, \sum_{i \in K} S_i = 1, 1 \leq i \leq K^A\}.$$

Для двох заданих зображень  $A$  і  $B$ , їх колірні характеристики є відповідно  $f_A = \{(I_i^A, S_i^A) \mid 1 \leq i \leq K^A\}$  та  $f_B = \{(I_j^B, S_j^B) \mid 1 \leq j \leq K^B\}$ . Колір  $I_k^B$ , з зображення  $B$ , який має мінімальну відстань до заданого кольору  $I_i^A$  визначається на основі евклідової відстані  $W$ :

$$k = \arg \min_{j \in K^B} W(I_i^A, I_j^B)$$

Знайдене  $k$  використовується, щоб обчислити міру відстані:

$$D[(I_i^A, S_i^A), f_B] = |S_i^A - S_k^B| \cdot W(I_i^A, I_k^B)$$

між  $i$ -м елементом ознаки  $(I_i^A, S_i^A)$  і ознакою  $f_B$ . Відносні розміри областей  $S$  використовуються як вагові коефіцієнти. Так, для кожного кольору в зображенні  $A$  знаходиться найближчий колір із зображення  $B$ .

#### 3.2. Сегментація складних об'єктів

Для забезпечення адаптивних властивостей САМ пропонується проводити вибір параметрів алгоритму для блоку сегментації складних об'єктів на основі їх розміру. Для виділення складних об'єктів розроблено алгоритм багатопорогової текстурної сегментації на основі математичного сподівання і дисперсії значень точок текстурного поля [5]. В якості текстурних ознак використано просторові моменти першого порядку. Результати застосування алгоритму для дослідження тканин молочної залози наведено у частині 5.

#### 3.3. Структурний аналіз гістологічних зображень

Для аналізу структурних патологічних змін у тканинах розроблено методи і алгоритми аналізу ГЗ на основі теорії кристалографічних груп [6]. Аналіз ГЗ складається із таких етапів: визначення відповідних точок на контурах

мікрооб'єктів, обчислення афінних перетворень між мікрооб'єктами, визначення групи симетрії для групи мікрооб'єктів (частини зображення).

Характерні точки визначаються на основі дискретних центральних моментів області: обчислюється кут нахилу головної осі, координати центру мас об'єкта. Для області визначаються три характерні контурні точки, на основі яких визначаються коефіцієнти афінних. Для оцінки ступеня структурної зміни тканини застосовується визначення групи симетрії для мікрооб'єктів. Кожна група симетрії повністю характеризується своїми породжуючими перетвореннями які можна записати у вигляді коду. Присвоїмо всім можливим перетворенням номери від 1 до  $l$ , а матрицям зсувів від 1 до  $k$ . Комбінацію власне перетворення і зсуву позначимо через верхній і нижній індекс відповідно  $T_k^l$ . Групу симетрії елементарної комірки опишемо через впорядковану послідовність породжуючих перетворень між першим елементарним рисунком та всіма іншими  $g^*(P) = (T_1^1, T_2^3, \dots, T_k^l)$ . Класифікація групи симетрії зводиться до послідовного співставлення послідовності породжуючих перетворень однієї з 17 груп  $g^*(P)$  із послідовністю утвореною для даного зображення  $g(P)$ . Визначення виду виміряного перетворення  $C$  елементарного рисунку проводиться шляхом його послідовного порівняння з усіма можливими еталонними перетвореннями  $T_j^i$ ,  $j=1, \dots, m$  для 17 груп симетрії площини і визначення найближчого згідно введеної відстані  $d$ . Відстань  $d(C, T)$  між матрицями двох перетворень  $C=(a_{ij})$  і  $T=(b_{ij})$  розраховується як сума відстаней між відповідними елементами матриць

$$d(C, T) = \sum_{i=1}^2 \sum_{j=1}^2 |a_{ij} - b_{ij}|.$$

Результатом аналізу ГЗ є його опис у вигляді матриць породжуючих перетворень. Матриці в числовому вигляді зберігаються у файл, що є описом даного ГЗ. За цим описом проводиться класифікація зображень, порівняння та пошук в БД.

#### 3.4. Синтез гістологічних зображень

Для наочності, сприйняття і передачі інформації а також формування навчальних вибірок зображень використовують синтез зображень. Синтез гістологічних зображень здійснено на основі використання плоских кристалографічних груп. В якості непохідного елемента (елементарного зображення) взято зображення клітини, над яким виконано породжуючі перетворення певної групи симетрії. Для моделювання структурних змін у тканинах здійснено параметричні спотворення породжуючих перетворень, в результаті чого отримано ряд структур: регулярні, близько-регулярні, нерегулярні, близько-стохастичні та стохастичні. Приклади порівняння алгоритмів синтезу приведено в роботі [7].

## 4. Технічна реалізація

Розроблене математичне та алгоритмічне забезпечення технології оброблення ГЗ використано при створенні програмного забезпечення САМ. Для написання програмного засобу використано програмне середовище IntelliJIDEA, мови програмування Java, Object Pascal, C++. Для роботи ПЗ необхідна робоча станція на базі операційної системи Windows. Програмний засіб розроблено як розширення (модуль) редактора ImageJ.

Для проведення аналізу вхідного зображення використано бібліотеку операцій програмного засобу ImageJ.

Розроблене ПЗ розв'язує такі задачі оброблення ГЗ: створення методики дослідження (сценарію), виконання дослідження в автоматичному режимі, робота з БД, робота з правилами діагностування. Засіб складається із наступних підсистем: редагування та запуск методики дослідження, редагування списку зображень дослідження, редагування правил діагностування. Модульна структура САМ зображена на рисунку 2.

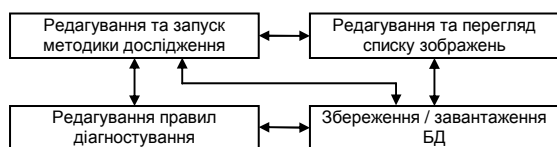


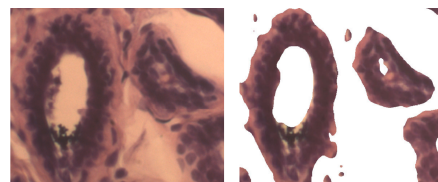
Рисунок 2: Структура програмного засобу аналізу ГЗ.

Для того, щоб зменшити вимоги до рівня знань в області комп'ютерного зору користувача було забезпечено можливість створення сценаріїв – методик дослідження, які використовуються для автоматизації часто повторюваних операцій. Методика створюється розробником САМ або користувачем. Готова методика, що використовує оцінку результатів операції оброблення зображення може здійснювати в автоматизованому режимі.

Типова структура апаратної частини САМ складається з системи вводу зображень (СВЗ), побудованої на базі мікроскопа, відеокамери або фотокамери, комп'ютера з програмним забезпеченням і принтера. СВЗ є складним компонентом, структура якого залежить від класу системи і вміщує в собі світловий мікроскоп, камеру, фотоадаптер, пристрій зміни поточного об'єктива, пристрій фокусування, пристрій переміщення предметного стола, пристрій подачі зразків та пристрій освітлення.

## 5. Експерименти

Для дослідження ефективності розробленого математичного та алгоритмічного забезпечення проведено його тестування системи аналізу ГЗ на базі цитологічних і гістологічних зображень Breast Cancer Dataset. Виділення дольок молочної залози з допомогою текстурної сегментації наведено на рисунку 3.



Вихідне зображення Результат сегментації

Рисунок 3: Детекція складних об'єктів

Розроблений алгоритм визначення відповідних точок на основі центральних моментів виявляється кращим при дії цілочисельної трансляції та повороту. Величина перекриття співставлених областей за допомогою розробленого алгоритму менша на 8% у порівнянні із методом січних.

## 6. Висновки

Програмне, математичне та алгоритмічне забезпечення для аналізу ГЗ розроблено в рамках тем Міністерства освіти і науки, молоді та спорту України "Розробка та дослідження методів аналізу зображень біомедичної природи" (2008-2012 рр.) та "Інформаційно-аналітична система для дослідження та діагностування пухлинних (ракових) клітин людини на основі аналізу їх зображень" (2008-2010 рр.). Розроблені методи та алгоритми дозволили підвищити достовірність діагностування злоякісних новоутворень, за рахунок кількісної оцінки патологічних змін в тканинах людини.

## 7. Література

- [1] Березький О.М., Березька К.М., Батько Ю.М., Мельник Г.М., "Синтез альтернативних рішень при структурному проектуванні систем автоматизованої мікроскопії," *Науковий вісник Українського державного лісотехнічного університету*, том. 23, №. 19(5), С. 258–268, 2009.
- [2] Абламейко С.В., Лагуновский Д.М., *Обработка изображений: технология, методы, применение. Учебное пособие.* Минск: Амалфея, 2000.
- [3] Томашевський О. М., Цегелик Г. Г., Вітер М. Б., Дудук В.І., *Інформаційні технології та моделювання бізнес-процесів.* К.: "Центр учбової літератури", 2012.
- [4] Berezsky O., Melnyk G., Batko Y., "Biomedical image search and retrieval algorithms," *Computing*, vol. 7, pp. 108–114, 2008.
- [5] Березький О.М., Мельник Г.М., Батько Ю.М., "Текстурна сегментація біомедичних зображень на основі просторових моментів," *Матеріали 4-ї Міжнародної "Комп'ютерні науки та інформаційні технології 2009"*, Л.: ПП "Вежа і Ко", 2009, С. 42–45.
- [6] Мельник Г.М., "Метод і алгоритми аналізу симетричних зображень," *Штучний інтелект*, том. 4, С. 253–261, 2010.
- [7] Березький О.М., Мельник Г.М., "Порівняння алгоритмів синтезу біомедичних зображень," *Інтелектуальні системи прийняття рішень і проблеми обчислювального інтелекту (ISDMCI 2011)*, 2011, С. 189–193.